

WHAT IS CLAIMED IS:

1. A system for classifying packets based on packet content, the system comprising:

5 a sequencer operable to receive packets and to identify packet flows;

a content engine interfaced with the sequencer to receive packets and to search packet contents for predetermined expressions in a packet or in a packet flow; and

10 a tag map interfaced with the content engine and operable to tag packets according to the predetermined expressions found by the content engine.

15 2. The system of Claim 1 wherein the sequencer comprises:

an enqueue engine operable to read packet flow sequencing information;

20 a packet flow tracker interfaced with the enqueue engine and operable to track packet flows with the sequencing information; and

a dequeue engine interfaced with the packet flow tracker and the content engine, the dequeue engine forwarding packets to the content engine according to sequencing information received from the content engine.

25

3. The system of Claim 2 wherein the enqueue engine is further operable to determine that a packet is out of order for that packet's flow and to transmit the out-of-order packet to have any missing packets resent.

30

4. The system of Claim 2 wherein the dequeue engine forwards the next packet of the flow for the sequencing information received from the content engine.

5 5. The system of Claim 4 wherein the dequeue engine determines that no packets for the packet flow are ready and determines a second packet flow to send to the content engine.

10 6. The system of Claim 1 wherein the content engine comprises:

a non-deterministic finite automata engine operable to search packet content for one or more regular expressions; and

15 one or more hash engines operable to search packet content for one or more subexpressions.

20 7. The system of Claim 6 further comprising a lexical analyzer interfaced with the non-deterministic finite automata engine and the hash engine, the lexical analyzer determining characters of the packets.

25 8. The system of Claim 6 wherein the non-deterministic finite automata engine comprises field programmable gate arrays.

9. The system of Claim 6 further comprising a state store module interfaced with the non-deterministic finite engine and operable to save the state of the non-deterministic finite automata engine associated with a 5 packet flow so that the saved state is available for the search of the next packet of the packet flow.

10. The system of Claim 6 further comprising a tag map interfaced with the content engine to map the packet 10 to a tag based on the content search.

11. A method for classifying packets based on content, the method comprising:

identifying packet flows;

searching packet content across the identified

5 packet flows to find one or more predetermined regular expressions;

computing a hash for predetermined strings of the regular expressions to find one or more subexpressions; and

10 tagging packets based on regular expression and subexpression matches.

12. The method of Claim 11 wherein the packet flow comprises a TCP stream.

15

13. The method of Claim 12 wherein identifying packet flows further comprises:

determining if a packet is out of order;

transmitting the out of order packet to its client

20 to have missing packets resent;

buffering the out-of-order packet until the missing packet is received; and

making the packet flow associated with the missing packet available for content searching.

25

14. The method of Claim 11 wherein searching further comprises finding regular expression matches by encoding the regular expressions as non-deterministic finite automata.

30

15. The method of Claim 14 further comprising:
computing a hash for a subexpression of a regular
expression match; and
finding a subexpression match if the computed hash
matches a hash in a hash look-up table.

16. The method of Claim 14 wherein the non-deterministic finite automata is encoded with field programmable gate arrays.

10

17. The method of Claim 16 further comprising:
storing the state of the field programmable gate arrays for a packet of a first packet flow;
searching the content of a packet of a second packet flow with the field programmable gate arrays;
loading the stored state of the first packet flow into the field programmable gate arrays; and
searching the next packet of the first packet flow.

18. A system for sequencing packet streams for content classification, the system comprising:

an enqueue engine that receives the streams and reads the stream identification of stream packets to
5 determine if a packet is out of order;

a stream tracker interfaced with the enqueue engine that associates packets to streams based upon the stream identification read by the enqueue engine;

10 a dequeue engine interfaced with the stream tracker and operable to forward packets for classification based on the packets' stream identification.

15 19. The system of Claim 18 further comprising a packet buffer interfaced with the enqueue engine for storing packets, the enqueue engine further operable to transmit an out-of-order packet, mark the out-of-order packet as sent and buffer the out-of-order packet so that any missing packet may be resent.

20 20. The system of Claim 18 wherein the dequeue engine is further operable to receive a stream identification, forward the next packet of the stream associated with the stream identification if the stream tracker indicates the next packet is ready, and forward
25 the next packet of a second stream if the next packet of the associated stream is not ready.

30 21. The system of Claim 20 wherein the dequeue engine updates the stream tracker to indicate when a packet is forwarded for classification.

22. A system for classifying packets based on packet content, the system comprising:

a lexical analyzer operable to determine the characters of a packet;

5 a content engine interfaced with the lexical analyzer and operable to determine if the packet characters match a predetermined expression;

a hash engine interfaced with the lexical analyzer and the content engine, the hash engine operable to 10 determine a hash for a subexpression of an expression match;

a hash look-up table interfaced with the hash engine to determine if the hash matches a predetermined string; and

15 a tag map that classifies the packet with a tag according to expression and subexpression matches.

23. The system of Claim 22 wherein the hash look-up table comprises a high bit engine that indexes the hash 20 by high bits and a low bit engine that matches low bits to strings determined by the high bit index.

24. The system of Claim 22 wherein the content engine comprises non-deterministic finite automata 25 encoded with field programmable gate arrays.

25. The system of Claim 22 further comprising a state store module interfaced with the content engine, the state store module storing the state of the content engine for a first packet of a first stream and loading 5 the content engine with the stored state when the next packet of the first stream is searched by the content engine.

26. The system of Claim 25 further comprising a 10 stream retriever that sends the first stream identification to request the next packet of the first stream and determines the stream identification of the received next packet to determine if the next packet is associated with the first stream.

15 27. The system of Claim 26 wherein the stream retriever instructs the state store module to store the first stream packet's state if the next packet received is associated with a second stream.

20 28. The system of Claim 22 wherein the content engine determines an expression match by completing plural states, with one state associated with each character of the expression.